Modeling the U.S. Processing Tomato Industry

Charles Plummer[1]

Abstract: The United States is the world's largest producer of processed tomato products. Over the last 20 years, average production has increased 47 percent and the farm value of the crop is now around $700 million. Processing tomatoes are second only to potatoes in terms of national per capita vegetable consumption. Given the significance of this industry, this article presents an econometric model designed to provide short run projections of this industry's key variables, such as acreage, yield, production, price, trade, and domestic use.

Keywords: Processing tomatoes, modeling, production, trade, prices, supply, demand.

The U.S. processing tomato industry has realized substantial growth and change over the past 20 years. In the late 1970's (1975-79), total processing tomato production averaged 7.3 million tons, with an average annual crop value of $463 million. The average crop size for the past 5 years (1994-98) was 10.7 million tons (47 percent above the late 70's) with an average annual crop value of $658 million (up 42 percent). The preliminary estimate for the 1999 crop is over 11.5 million tons, just shy of the record set in 1994, and the crop value will likely top $700 million.

With such tremendous growth in the industry over the past two decades, older econometric forecasting models of the processing tomato industry are unable to capture more recent structural and demand changes, and are therefore unlikely to yield useful forecasts. The purpose of this article is to present an econometric model designed to provide short run projections of the industry's key variables, such as acreage, yield, production, price, trade, and domestic use. An overview of industry structure and domestic and export demand is followed by development of the model and model forecasts.

Industry Overview

Much of the increase in tomato production over the past 20-30 years can be attributed to improved yields and increased efficiency at the grower and processor levels. In the late 1970's (1975-79), the average yield of processing tomatoes in the United States was 22.1 tons/acre. By the late 1980's the average had risen to 28.5 tons/acre, and in the late 1990's the average has been 33.2 tons/acre. This increase in yields is due to the steady development of higher-yielding hybrid varieties, improved cultural practices such as increased use of transplanting, and a continued shift in production to California, where average yields are currently about 30 percent higher than the average yield in the rest of the United States. In the late 1970's, California accounted for about 84 percent of total U.S. processing tomato production. At the end of the 1990's, California is averaging approximately 94 percent of the total U.S. production.

In addition to improving raw tomato yields, hybrid varieties (along with advances in mechanization) have helped to improve final product output. Hybrid tomato characteristics (such as skin thickness, solids content, etc.) have helped harvesting equipment to become more efficient, and have helped improve processor recovery rates and product quality. Harvest has become less labor intensive and processors are now more able to meet stringent output standards for final product.

These improvements have resulted in some overall structural change within the industry. Several major companies have restructured to source a large portion of their product needs through other tomato processors who can meet their quality standards. This restructuring has helped to improve overall industry efficiency, as processing capacity consolidates and older, less efficient processing plants are taken out of production (Welty). For example, 1992 California tomato paste processing capacity (tomato paste output) from direct marketers (i.e., not including remanufacturers) was 484,000 pounds/hour from a total of 16 processing facilities. By 1998, total capacity had risen to 698,000 pounds/hour from a total of 13 facilities. Additionally, of the 16 facilities in operation in 1992, 8 had been built prior to 1976. By 1998, the number of facilities in use built before 1976 had decreased to five (The Food Institute).

Growth in Domestic Demand

Tomatoes and tomato products are an important part of the American diet. After potatoes, tomatoes are the most widely consumed vegetable in the United States. During the past 20 years, U.S. annual per capita use of tomatoes and tomato

[1] Agricultural economist, Market and Trade Economics Division, Economic Research Service, U.S. Department of Agriculture.

products has increased by nearly 30 percent to a total fresh-weight equivalent of 93 pounds/person in 1998. Processed tomato products, including sauces, ketchup, pastes, salsa, juice, etc., accounted for 81 percent of that total. Utilization statistics by processed product type are unavailable. However, an estimate based on limited data from the late 1980's suggests that the largest use is for sauces (35 percent), followed by paste (18 percent), canned tomatoes (17 percent), ketchup, and juice (each about 15 percent).

Domestic use of processed tomato products has increased substantially in the 1990's. During the 1980's, per capita use of processed tomato products averaged 63.5 pounds. During the 1990's, processed use has averaged 75.5 pounds per person. The increase is likely the result of continued expansion in food-service demand, especially for pizza, tacos, and other Italian and Mexican foods.

Some of the increase may also be due to rising public awareness of the health benefits of processed tomato products in the diet. Several studies this decade have linked diets rich in tomatoes and tomato products to reduced risk of various cancers and heart disease. Tomatoes contain lycopene, a naturally occurring compound that, when ingested, acts as a powerful anti-oxidant that helps protect human cells from the degenerative effects of various free radicals. Some studies indicate that canned tomato products may contain higher concentrations of lycopene than fresh tomatoes due to the heat processing used to create the product.

Although domestic consumption of processed tomato products has boomed in the 1990's, it appears that trend may be leveling somewhat as the century closes. Per capita use has averaged just under 75 pounds for the past 5 years, with the preliminary estimate for 1999 at 75.3. However, with continuing strong export potential in the coming decade, slow growth (or even a slight decline) in domestic demand in the next few years does not necessarily translate into little or no growth in domestic production.

Exports Could Drive Future Production

Even with domestic consumption of processed tomato products appearing to level somewhat, the outlook for continued growth in U.S. production looks good as there appears to still be potential for continuing growth in exports. The United States is the world's largest producer of processed tomato products, and exports have just recently become an increasingly important outlet for U.S. producers. Prior to 1989, exports of processed tomato products rarely accounted for more than 1 to 2 percent of total processed tomato utilization (on a raw-equivalent basis). However, since 1989, the importance of exports has steadily risen, and in 1998 exports accounted for 12 percent of total utilization. In 1989, all U.S. processed tomato product exports were valued at $60.1 million. By 1998, total value had nearly quadrupled, rising to $237.1 million.

Such a dramatic increase can be partly accredited to increased access to the Canadian market as a result of the North American Free Trade Agreement (NAFTA). In 1989, Canada accounted for 25 percent ($15.1 million) of total export value of all processed tomato products. By 1998, Canada's take had risen to 50 percent ($119.3 million). NAFTA has also increased access to the Mexican market, although exports to Mexico are still relatively small (accounting for 6 percent of total export value in 1998). Exports to Mexico increased from $3.4 million in 1989 to $14.4 million in 1998.

Exports of processed tomato products have also increased to other regions of the world, although the growth rate has not been as rapid as it has been to Canada and Mexico. Export value to Japan, the second largest U.S. export market for processed tomato products, rose 251 percent from 1989 to 1998. Japan accounted for $27.3 million (12 percent) in processed tomato exports in 1998, of which, 22 percent was ketchup. Other Asian and Pacific Rim countries took nearly $23 million in U.S. exports in 1998, with nearly 29 percent of that being ketchup. The rapid growth of Western-style fast food chains in Japan and other Asian Pacific Rim countries over the last decade has spurred the growth in U.S. processed tomato exports to the region. As Western-style cuisine continues to increase in popularity around the world, the United States should remain well situated to continue increasing exports of processed tomato products.

Model Development

An econometric model was developed to estimate U.S. processing tomato acreage, beginning stocks, production, grower price, trade, and domestic utilization. The system of equations can be found in table A-1. The supply portion of the model consists of an equation to estimate beginning stocks, an acreage response function, a yield equation, a multiplicative identity for production, and an import equation. The demand portion of the model is determined by an export equation, with domestic utilization being the residual (beginning stocks + domestic production + imports – ending stocks – exports). Because nearly 100 percent of anticipated production is contracted between processors and growers prior to planting, at an agreed upon price, acres planted is the best variable (when combined with an autoregressive [AR] term) in forecasting grower price. The model is estimated using annual data from 1970 to 1998. Endogenous and exogenous variables are identified in table A-2.

The method used to minimize the annual variation of an endogenous variable is ordinary least squares (OLS) regression. The explanatory variables are either economic variables logically consistent with microeconomic theory, proxy variables such as trend, or related forecast variables such as Economic Research Service (ERS) or National Agricultural Statistics Service (NASS) estimates. In all but two of the regression equations, first-order serial correlation appeared to be a significant problem (as indicated by a Durbin-Watson statistic below the critical value at the 5 percent level). In these equations, autoregressive (AR) and/or moving average (MA) terms were added to account for the problem. When current endogenous variables appear as independent variables (right-hand side variables) in regression equations, OLS may produce biased and/or inconsistent parameter estimates. However, since these problems didn't appear to be severe in any of the equations, all published equations are OLS regression results.[2] Selected in-sample summary statistics are found in table A-3.

Forecasting Assumptions and Results

For any model to forecast successfully, some assumptions are necessary. First, the model must be correctly specified to reflect market behavior. The in-sample summary statistics (table A-3) partially confirm this assumption. Additionally, it must be assumed that the behavior captured by the model must continue into the future. Naturally, this assumption weakens as the forecasting horizon extends because the model is unable to capture out-of-sample (future) dynamics without being re-estimated. Finally, several exogenous factors must be given future values for estimation. For this model, the exogenous variables future values are forecast using AR and ARMA models. Thus, caution must be used when interpreting the model's forecasts, as they are based on a set of explicit assumptions that include forecasts of determinate variables. Changes in these assumptions can have significant effects on the model's forecasts. The equations used to determine future values of the exogenous variables are in table A-1 and selected in-sample summary statistics can be found in table A-3.

The best test of a forecasting model's accuracy is a comparison of forecasts with actual values. Because recent structural changes have had such an impact on the processing tomato industry, the model was estimated using data through 1998 in order to capture these changes in the explanatory variables. This currently leaves 1999 as the only out-of-sample year for which to evaluate the model's accuracy. Preliminary USDA estimates of 1999 processing tomato acreage, yield, and production indicate fairly good agreement between model forecasts and actual (preliminary) values (table A-4).[3]

For the 2000 simulation, the model was run using USDA's July and September preliminary estimates for 1999 values of contract acreage, yield, and production. Because 1999 seems to be shaping up as a somewhat unusual trade year, trade volume for 1999 was not estimated by model equations. Export volume in 1999 was estimated to be 15 percent below 1998 levels, and import volume was estimated to be 20 percent above 1998. Model equations were used to estimate the remaining 1999 variables and all year 2000 estimates. The results of this simulation indicate declines of 9

[2] Two-stage least squares regressions were run on relevant equations and the results were not significantly different from OLS results. Therefore, all published equations are OLS regression results.

[3] Current USDA estimates are for contracted production only. However, this typically accounts for 99 percent of the industry total, which will be published in January 2000.

Table A-1--Regression equations for U.S. processing tomato mode

BSTOCKS	= -3138464 + .667947(Q(-1)+BSTOCKS(-1)) - 1.517443(EXPVOL(-1)) + .659246(MA(1)
	(-5.171) (11.602) (-3.139) (4.608)
ACPL	= -682332.5 + 61250.26(LOG(EXPVOL(-1))) - .021608(BSTOCKS) + 256785.3(CNPPI(-1)/PPITW(-1)) + .442659(AR(1))
	(-4.682) (6.013) (-4.057) (5.927) (2.440)
ACHV	= .969680(ACPL)
	(245.115)
YD	= -1043.467 + .539305(TIME)
	(-17.310) (17.750)
Q	= (ACHV)(YD)
GRP	= 35.12016 + .0000904(ACPL) + .813855(AR(1))
	(3.843) (3.595) (9.857)
IMPVOL	= 94046.65 + 6776.403(CNPPI) - .095740(BSTOCKS) + .769664(AR(1)) - .497805(AR(2)) - .954630(MA(1))
	(3.317) (18.667) (-12.800) (4.010) (-3.514) (-48.543)
LOG(EXPVOL)	= 17.25270 + .977480(AR(1))
	(1.352) (17.377)
CNPPI	= 134.9139 + .897121(AR(1)) + .547769(MA(1))
	(5.405) (14.904) (3.194)
PPITW	= 145.2909 + .953036(AR(1)) + .532149(MA(1))
	(3.545) (35.112) (3.192)

Table A-2--Variable definitions

Exogenous:		
CNPPI	=	U.S. producer price index for canning tomatoes (index 1982 = 100)
PPITW	=	U.S. prices paid for interest, taxes, and wages (index 1990-92 = 100)
TIME	=	Year (1970-1998).
Endogenous:		
BSTOCKS	=	U.S. stocks of processed tomato products on Jan. 1 (short tons).
ACPL	=	U.S. total annual planted acreage of processing tomatoes (acres)
ACHV	=	U.S. total annual harvested acreage of processing tomatoes (acres)
YD	=	U.S. total annual processing tomato yield per acre harvested (tons/acre)
Q	=	U.S. total annual processing tomato production (tons)
GRP	=	U.S. total average grower price for processing tomatoes ($/ton)
IMPVOL	=	U.S. total annual import volume of processed tomato products (tons)
EXPVOL	=	U.S. total annual export volume of processed tomato products (tons)

Table A-3--Selected in-summary statistics, 1970-98

Equation	Adjusted R-squared	Standard error of regression	Durbin-Watson statistic
BSTOCKS	0.962	327817.700	2.500
ACPL	0.750	19111.820	1.920
ACHV	0.968	6649.586	2.402
YD	0.918	1.369	1.430
GRP	0.812	4.435	2.144
IMPVOL	0.802	80627.250	2.106
LOG(EXPVOL)	0.918	0.287	1.484
CNPPI	0.956	6.312	1.795
PPITW	0.991	2.406	1.908

percent in acreage, 6 percent in production (assuming trend yields), and 4 percent in average grower prices (table A-4).

The model forecasts a 9-percent decline in acreage next year although other market factors seem to indicate that a smaller decline is possible. The 2000 season looks now as though it may have a beginning very similar to 1995. It looks as though beginning stocks in the year 2000, although significantly higher than 1999, will not be substantially higher than levels experienced recently in the industry. As was the case in 1995, the 2000 season will follow 2 previous years of declining stocks, and one year of relatively high production. In 1995, growers increased acreage under these similar circumstances, and production was relatively high for a second consecutive year. This, subsequently, contributed to increased beginning stocks again for 1996. Due to the relatively large stocks on hand, acreage was decreased in 1996. However, strong yields led to increased production and record-high beginning stocks in 1997. This forced a large drop in acreage and production in 1997 followed by only a slight acreage increase in 1998. During this 3-year adjustment period (1996-1998), grower prices decreased each year.

In order to avoid the potential of a similar build-up in stocks over the next several years, it is likely that acreage will decrease in 2000, but perhaps not quite the 9-percent forecast by the model. With increased grower prices in 1999, and a good export outlook in the coming years, a more gradual decline in acreage is more likely. If planted acreage decreased about 5 percent to 335,000 acres in 2000, average acreage abandonment and trend yields (which would approach record-high levels) would lead to only a slight decline in production.[4] If only a slight decline in production is realized next year, stocks would likely rise substantially in 2001 and trigger a larger decline in acreage. The forecasts for the various scenarios can be seen in table A-4.

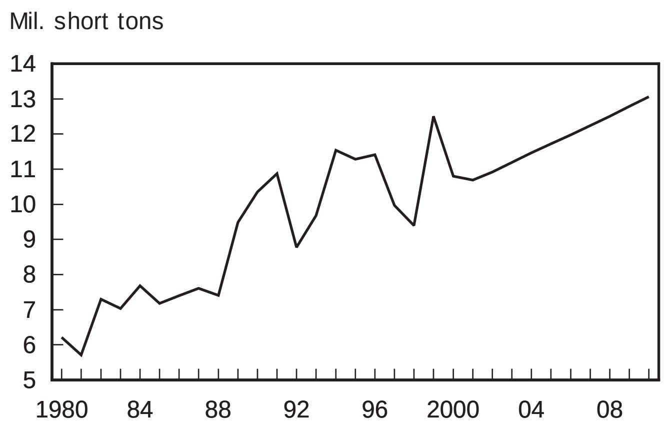
Long-run Outlook

Long-run forecasts (from 2001 and beyond) with the model are somewhat impractical, as the potential for forecasting error increases with the forecast horizon. Over time, current assumptions and the relationships estimated by the model will likely change and new relationships, not captured by the existing model, may become important. However, the model can be used as a reference point to speculate on long-run prospectus in the industry. For example, the model indicates that per capita domestic use of processed tomato products may decline slightly over the next couple of years and then increase slowly (table A-5). Likewise, production is forecast to decline for the next 2 years before steadily increasing to nearly 13.1 million tons in 2010 (table A-4).

[4] The 95 percent confidence interval for ACPL in the year 2000 is 277,250 to 356,526 acres.

Figure A-1

Processing tomato production



Sources: National Agricultural Statistical Service and Economic Research Service, USDA.

Table A-4--Actual and forecast values for processing tomatoes, 1998-2010 1/

Year	BSTOCKS	ACPL	ACHV	YD	Q	GRP	EXPVOL	IMPVOL
	Short tons	Acres	Acres	Tons/acre	Short tons	$/ton	Short tons	Short tons
1998	7,256,802	302,260	299,760	31.34	9,394,810	57.70	1,238,092	465,540
1999 p	--	352,860	347,750	33.18	11,538,066	--	--	--
1999 f	5,670,019	339,113	328,831	34.60	11,379,130	61.92	1,331,315	374,004
1999 fa	5,670,019	352,860	347,750	33.18	11,538,066	63.16	1,052,378	558,647
2000 f	6,444,102	316,881	307,280	35.14	10,799,067	60.63	1,135,766	174,628
2000 fa	6,444,102	335,000	324,843	35.14	11,416,295	62.26	1,135,766	224,720
2001 f	7,067,877	308,937	299,570	35.68	10,689,665	60.49	1,223,659	67,481
2002 f	6,865,803	310,802	301,379	36.22	10,916,739	61.14	1,316,142	194,604
2003 f	6,742,163	314,046	304,524	36.76	11,194,907	61.82	1,413,295	341,641
2004 f	6,697,957	316,881	307,274	37.30	11,461,697	62.39	1,515,186	399,602
2005 f	6,692,017	319,422	309,737	37.84	11,720,640	62.88	1,621,879	377,200
2006 f	6,699,109	321,875	312,116	38.38	11,978,958	63.31	1,733,426	334,769
2007 f	6,707,122	324,355	314,520	38.92	12,240,882	63.70	1,849,872	315,700
2008 f	6,710,726	326,917	317,004	39.46	12,508,521	64.07	1,971,253	324,205
2009 f	6,707,714	329,583	319,590	40.00	12,782,910	64.43	2,097,593	342,346
2010 f	6,697,265	332,364	322,287	40.54	13,064,585	64.77	2,228,911	354,413

-- = not available.

1/ 1998 are actual values. "p" denotes preliminary estimates from NASS, USDA. "f" denotes model forecasts. "fa" denotes model forecasts adjusted using other current information. Forecasts from 2001-2010 run from 2000fa values.

Table A-5--Model projections of domestic use, 1998-2009 1/

	Total domestic use	Per capita use
	Million lbs.	Pounds
1998	20,418	75.6
1999	20,541	75.4
2000	19,763	72.0
2001	19,471	70.3
2002	19,838	71.1
2003	20,335	72.3
2004	20,704	73.0
2005	20,938	73.2
2006	21,145	73.4
2007	21,406	73.7
2008	21,729	74.2
2009	22,076	74.8

Conclusions

As the model suggests, exports (EXPVOL) and cost-effective (CNPPI[-1]/PPITW[-1]) production are key variables in determining planted acreage. Continued growth in export markets through a rapidly expanding food-service industry and increasing consumer awareness about the effects of a healthy diet should both bode well for processing tomato demand in the years to come. If the U.S. producers can remain relatively low-cost producers of high-quality processed tomato products, the industry should be well situated to capitalize on increasing worldwide demand. Conversely, slow growth in exports and/or increasing producer costs relative to returns would likely translate into little or possibly even negative growth in the domestic industry.

Because beginning stocks, exports, and producer costs relative to returns are such important variables in forecasting planted acreage, reliable forecasts of these variables are vital to forecasting acreage, production, and price. Perhaps the most glaring weakness of this particular forecasting model is the lack of a theoretically satisfactory export volume equation. While the simple AR process fits well statistically and accounts for the relatively recent explosion in exports, it does nothing to account for potential changes in key variables which would likely affect export volume, such as changes in price, exchange rates, and stocks on hand. An improvement in long-run export forecasting would likely improve the model's other long-run projections.

References

Chern, W.S., and R.E. Just. "Econometric Analysis of Supply Response and Demand for Processing Tomatoes in California." University of California, Berkley Agricultural Experiment Station. Giannini Foundation Monograph No. 37, 1978.

Conway, Roger K. and John F. Yanagida. "Estimating U.S. Potato Demand: Structure and Forecasts." USDA, ERS. *Vegetables and Specialties Situation and Outlook Report,* TVS-218 (1980): 31-40.

The Food Institute. "Food Markets in Review, 1998/99: Tomato Products." Fair Lawn, NJ, 1998, pp. 2-5.

Love, John M. and Lois S. Willet. "Modeling the U.S. Potato Industry." USDA, ERS. *Vegetables and Specialties Situation and Outlook Report*, TVS-250, March 1990, pp. 16-24.

Welty, John C. "The California Tomato Growers Association's District Meeting Wrap-Up." California Tomato Growers, Inc.